日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: １９９９年１２月　９日

出願番号
Application Number: 平成１１年特許願第３４９８０９号

出願人
Applicant (s): 甲府日本電気株式会社

２０００年　９月２２日

特許庁長官
Commissioner,
Patent Office 及川耕造

出証番号　出証特２０００－３０７８１５１

【書類名】　特許願

【整理番号】　03905037

【提出日】　平成11年12月 9日

【あて先】　特許庁長官殿

【国際特許分類】　G06F 15/00

【発明者】

【住所又は居所】　山梨県甲府市大津町１０８８－３　甲府日本電気株式会社内

【氏名】　近藤　秀俊

【特許出願人】

【識別番号】　000168285

【氏名又は名称】　甲府日本電気株式会社

【代表者】　桑田　幹雄

【代理人】

【識別番号】　100105810

【弁理士】

【氏名又は名称】　根本　宏

【手数料の表示】

【予納台帳番号】　072627

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9912421

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ネットワークシステムにおけるデータアクセス方法、ネットワークシステムおよび記録媒体

【特許請求の範囲】

【請求項１】　複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項２】　複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりも読み出し速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一

のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項３】　複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項４】　複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備える共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリか

らデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステム。

【請求項５】　複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備えると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一データがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステム。

【請求項６】　請求項４および５のいずれかに記載のネットワークシステムにおいて、

前記タグメモリを前記通信機構に設けたことを特徴とするネットワークシステ

ム。

【請求項７】　複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法を実現するためのデータアクセス用プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスする処理と、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する処理と、を含む処理をコンピュータに実行させるためのデータアクセス用プログラムを記録した記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、ネットワークシステム、特にＮＵＭＡ（Non Uniform Memory Access model）システムに適用して好適なデータアクセス技術に関する。

【０００２】

【従来の技術】

図２１は、ＮＵＭＡシステムの一例のブロック構成図であり、このシステムでは、複数のノード装置１０００ａ、…、１０００ｎの夫々が、互いに所要の情報を通信可能にバス１０４０で接続されていて、各ノード装置は、ＣＰＵ１０１０ａ（…、１０１０ｎ）と、ローカルメモリ１０２０ａ（…、１０２０ｎ）と、制御ユニット１０３０ａ（…、１０３０ｎ）とを有している。また、図２２に示すように、各ノード装置のローカルメモリ１０２０ａ、…，１０２０ｎは、共有メモリを構成していて、各ノード装置のＣＰＵ１０１０ａ、…、１０１０ｎは、この共有メモリをアクセス可能に構成されている。

【０００３】

そして、このようなNUMAシステムにおいては、各ノード装置１００ａ、…，１００ｎ内のＣＰＵ１０１０ａ、…、１０１０ｎから共有メモリへアクセス動作を行なう場合、読み出し開始からデータ読み出しが完了するまでの期間であるレイテンシが異なってしまうため、ＣＰＵ１０１０ａ、…、１０１０ｎの夫々が、キャッシュを内蔵することで、自ノード装置（以下、ローカルノードと称する場合もある）が、他ノード装置（以下、リモートノードと称する場合もある）のＣＰＵ内のキャッシュへアクセスする回数を減らすように工夫を施している。

【発明が解決しようとする課題】

しかしながら、共有メモリに記憶されているデータをキャッシュメモリにコピーする際に特に制約がない場合にあっては、共有メモリにおける同一アドレスのデータコピーが、複数のプロセッサ内のキャッシュメモリ内に保持されてしまい、ノード装置間におけるアクセスが頻繁に発生する。したがって、リモートノードのキャッシュメモリやローカルメモリに対するアクセス時のレイテンシ悪化がシステム性能低下をもたらしていた。

【０００４】

また、このシステムにあっては、共有メモルを構成する各ローカルメモリが、ノード装置間に分割して実装されるため、ローカルノードとリモートノードでのノード間のレイテンシ差が大きくなり、これもシステム性能低下をもたらしていた。

【０００５】

本発明は、このような従来の課題を鑑みてなされたもので、レイテンシを短縮できるデータアクセス方法、ネットワークシステム、データアクセス用プログラムを記録した記録媒体を提供することを目的としている。また、本発明の他の目的は、ノード間のレイテンシ差が小さくすることにある。

【課題を解決するための手段】

上記目的を達成するために、本発明の内の請求項１に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に

構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定することを特徴とした。

【０００６】

この請求項１に係る発明ににおいては、タグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、読み出したタグ情報に基づいて、投機アクセスによりメモリから獲得されたデータを廃棄するか否かを決定する。したがって、従来のように、まず、タグ情報を読み出してこの読み出したタグ情報が、システム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、さらに、メモリアクセス動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行してメモリアクセスを行なっておいて、読み出したタグ情報がシステム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、既にアクセスしておいたデータを採用するので、レイテンシを短縮することができる。

【０００７】

また、請求項２に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりも読み出し速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法である。

【０００８】

この請求項２に係る発明においては、各ノード装置は、タグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送る。一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【０００９】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、自ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行して自ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出しをしておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【００１０】

また、請求項３に係る発明は、複数のノード装置の夫々が互いに所要の情報を

通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法である。

【００１１】

この請求項３に係る発明によれば、各ノード装置は、タグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【００１２】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを

示す場合には、さらに、他ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行して他ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【００１３】

また、請求項４に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備える共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステムである。

【００１４】

この請求項４に係る発明においては、投機読み出し手段は、タグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機的に読み出す。そして、読み出しデータ処理手段は、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても

存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【００１５】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、自ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行して自ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【００１６】

また、請求項５に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備えると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする投機読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、

このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステムである。

【００１７】

この請求項５に係る発明においては、投機読み出し手段は、タグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする。そして、読み出しデータ処理手段は、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【００１８】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、他ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行して他ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【００１９】

また、請求項６に係る発明は、請求項４および５のいずれかに記載のネットワークシステムにおいて、

前記タグメモリを前記通信機構に設けたことを特徴とする。

【００２０】

この請求項６に係る発明においては、上記効果に加えて、タグメモリを通信機構に設けることによって、いずれのノード装置からのタグ読み出しも略等しい時

間で行なえるので、ノード装置間のレイテンシ差を小さくすることができる。

【００２１】

また、請求項７に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法を実現するためのデータアクセス用プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスする処理と、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する処理と、を含む処理をコンピュータに実行させるためのデータアクセス用プログラムを記録した記録媒体である。

【００２２】

この請求項７に係る発明においても、タグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、読み出したタグ情報に基づいて、投機アクセスによりメモリから獲得されたデータを廃棄するか否かを決定する。したがって、従来のように、まず、タグ情報を読み出してこの読み出したタグ情報が、システム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、さらに、メモリアクセス動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行してメモリアクセスを行なっておいて、読み出したタグ情報がシステム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、既にアクセスしておいたデータを採用するので、レイテンシを短縮することができる。

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照しつつ説明する。以下、本発明の理解の容易化のため２台のノード装置でＮＵＭＡシステムを構成する場合を例にとっ

て説明するが、３台以上のノード装置でＮＵＭＡシステムを構成するようにしても良い。図１に示すネットワークシステムは、２台のノード装置１００ａ、１００ｂを相互に所要の情報を通信可能にＩＣＮ（Internal Connection Network：インターナルコネクションネットワーク）３００で接続している。

【００２３】

両ノード装置１００ａ、１００ｂは同一の構成であるので、ノード装置１００ａ側のみの構成を説明すると、このノード装置１００ａは、メモリユニット（ＭＵ）３０ａと、このメモリユニットよりもアクセス速度が高速なキャッシュメモリ２０ａを内蔵する１台のプロセッサ１０ａと、アクセス制御等を行なうシステムコントロールユニット（ＳＣＵ）２００ａとを備えていて、プロセッサ１０ａは、ライトバックやライトスルー等の公知のアルゴリズムによってキャッシュコヒーレンシを保証するように構成されている。

【００２４】

システムコントロールユニット（ＳＣＵ）２００ａは、プロセッサ１０ａとの間でインターフェイス制御動作を行うＰＩＵ（Processor Interface Unit：プロセッサインターフェイスユニット）４０ａと、ＭＵ（Memory Unit：メモリユニット）３０ａとの間でインターフェイス制御動作を行うＭＩＵ（Memory Interface Unit：メモリインターフェイスユニット）３５ａと、他ノード装置１００ｂと間でインターフェイス制御動作を行うＳＩＵ（System Interface Unit：システムインターフェイスユニット）４５ａとを有していて、ＰＩＵ４０ａは、プロセッサ１０ａから発行されるトランザクションのアドレスに基づいて、ホームノード（このアドレスが存在するノード装置）を検出する機能を有する。そして、図４に示すように、ノード装置１００ａのメモリユニット（ＭＵ）３０ａと、ノード装置１００ｂのメモリユニット（ＭＵ）３０ｂとで共有メモリを構成していて、両ノード装置１００ａ、１００ｂのプロセッサ１０ａ、１０ｂは、この共有メモリをアクセス可能に構成されている。

【００２５】

通信機構としてのＩＣＮ３００は、データのルーティングを行うためのルーティング部３２０と、タグ情報を記憶するタグメモリ３１０とを備えている。図２

に示すように、タグ情報は、ブロックデータの番号（ブロック番号）と、ステータス情報と、ノード装置番号とが対となっている。図３に示すように、ステータス情報「U」は、いずれのノード装置１００ａ、１００ｂ内のキャッシュメモリ２０ａ、２０ｂにおいても、アクセス対象となるデータが無いことを示すもので、この場合ノード装置番号は与えられない。ステータス情報「Ｓ」は、キャッシュメモリ内のデータはメモリユニット内のデータと一致しており、かつ、複数のノード装置（この場合、例えば１００ａ、１００ｂ）のキャッシュメモリ２０ａ、２０ｂにおいてアクセス対象となるデータと同一のデータが記憶されていることを示すもので、この場合には当該キャッシュデータを持つノード装置の番号がノード装置番号として対応付けられる。そして、ステータス情報「Ｐ」は、１つのノード装置（この場合、１００ａまたは１００ｂ）にのみ、キャッシュデータが有ることを示すもので、この場合には、このキャッシュデータを持つノード装置の番号がノード装置番号として対応付けられる。

【００２６】

図５は、本発明の主要部の構成図である。ＰＩＵ４０ａは、プロセッサ１０ａに送るリプライデータを複数エントリ分格納するＲＤＢ（Reply Data Buffer：リプライデータバッファ）４２０と、ＲＤＢ４２０のエントリ数分だけ設けられた制御ユニット（Ｗ１～Ｗｎ）が設けられたＲＤＣＢ（Reply Data Control Buffer：リプライデータコントロールバッファ）４１０と、ＲＤＣＢ４１０から、プロセッサ１０ａへのリプライが許可されたエントリーのうちから、いずれか１つのエントリを調停出力してＲＤＢ４２０に格納させるリプライ調停部４３０とを備えている。

【００２７】

図６に示すように、ＲＤＣＢ４１０を構成する一つの制御ユニット（例えばW１）は、ビットメモリ４１１と、Ｖビット生成部４１２とを備えていて、エントリ単位に設けられる。図７に示すように、ビットメモリ４１１に格納されるビットは６種類ある。「ＴＶビット（Transaction V bit）」は、ＰＩＵ４０ａがプロセッサ１０ａからのリード系トランザクションを受け付け、この受け付けたリード系トランザクションをＭＩＵ、ＳＩＵへ発行したことを示すビット、「ＬＶ

ビット（Local Memory V bit ）」は、ローカルノード内のメモリユニット（ＭＵ）であるローカルメモリからリプライデータが返却されたことを示すビット（Local Memory V bit）、「HVビット（Home Node V bit）」は、他ノード装置がホームノードであった場合、ホームノードのメモリユニット（ＭＵ）からリプライデータが返却されたことを示すビット、「ＣＶビット（Cache V bit）」は、リモートノード内のキャシュメモリからリプライデータが返却されたことを示すビット、「ＪＶビット（Judgement V bit）」は、ＳＩＵから返却されるスヌープ結果を示すビット、「EVビット（Entry V bit）」は、ＬＶ、ＴＶ、ＨＶ、ＣＶ、ＪＶビットに基づいて、対応するエントリーのＲＤＢ４２０に格納するリプライデータをプロセッサ１０ａへ送信可能であることを示すビットである。なお、タグメモリ３１０からタグ情報を読み出す動作を「スヌープ動作」と称し、この読み出し結果を「スヌープ結果」と称している。

【００２８】

Ｖビット生成部４１２は、ＬＶ、ＴＶ、ＨＶ、ＣＶ、ＪＶビットに基づいてEVビットを生成する。しかし、このＥＶビット生成がＰＩＵ４０ａへ報告されるタイミングは、各エントリー間で調整されていないため、複数エントリに対して複数のＥＶビットが立った場合、リプライ調停部４３０が調停出力した、ＲＤＢ５０８のエントリーに格納されたリプライデータが、プロセッサ１０ａに送られると共に、Ｖビット生成部４１２は、ＲＤＣＢ５０９のＥＶビットをリセットしてエントリの無効化を行うように構成されている。一方、スヌープ処理と並行して、ＣＰＵ１０ａから受け付けたリード系トランザクションに基づいて、ローカルメモリであるＭＵ３０ａやリモートノード１００ｂ内のＭＵ３０ｂを投機アクセスする。図５に示す例では、ローカルメモリであるＭＵ３０ａを投機読み出しするための専用パスである投機リードパス５００によって、「ＰＩＵ４０ａ→ＭＩＵ３５ａ→ＭＵ３０ａ」の経路でローカルメモリであるＭＵ３０ａからのデータ読み出しを行う。

【００２９】

ＳＩＵ４５ａは、ＰＩＵ４０ａから発行されたリード系トランザクションに応答して行なわれるタグメモリ３１０へのスヌープ処理のスヌープ結果待ちデータ

や読み出したスヌープ結果を保持するＳＣＢ（Snoop Control Buffer：スヌープコントロールバッファ）４５０と、ＳＣＢ４５０からスヌープ結果を読み出し、各トランザクションのスヌープ結果を生成して、これを判定情報としてＲＤＣＢ４１０へ出力する判定部４６０とを備えている。図８、図９に示すように、ＳＣＢ４５０には、エントリー数分の２種類の情報を格納する。この２種類の情報は、ＰＩＵ４０ａから発行されたリード系トランザクションに応答して実行すべきスヌープ処理が待ち状態であることを示す「Vビット」と、タグメモリ３１０からのスヌープ処理結果を表す「ＳＮＰ情報」である。Ｖビットが立った状態でスヌープ結果が得られると、判定部４６０によって、対象トランザクションのＲＤＣＢ４１０内のエントリに対応する判定情報を発行し、ＪＶビットの設定を行う。なお、判定部４６０で判定されたＳＣＢ４５０内のエントリのVビット、ＳＮＰ情報をクリアして、ＳＩＵ４５ａ内のトランザクション処理を完了するように構成されている。

【００３０】

また、判定部４６０は、タグメモリ３１０のスヌープ結果に基づいて判定情報を生成、出力するが、その判定情報は、キャシュコヒーレンシのプロトコルに依存する。例えば、一般的な「ＭＥＳＩ」プロトコルを用いた場合、ＲＤＣＢ５０９に送る判定情報としては、リモートノード１００ｂのキャシュメモリ２０ｂがModify Lineでデータを保持していることを示す「Modify有」（キャッシュメモリにデータが有る）と、リモートノードのキャシュがModify Lineでデータ保持していないことを示す「Modify無」（キャシュメモリにデータが無い）の２種類となる。

【００３１】

（第１の動作例）

次に動作説明をする。まず、図１０のケース①、即ち、ノード装置１００ａをホームノードとし、ローカルメモリ（ＭＵ３０ａ）に対するリード系トランザクションが、プロセッサ１０ａから発行された場合について、図１１等を参照しつつ説明する。なお、この時、図１５に示すようにタグメモリ３１０に記憶されるブロック番号ａのブロックデータに属するいずれかのデータをリード系トランザ

クションの読み出し対象とする。

【００３２】

図１１の下側には、このケースの動作の処理手順を示すタイムチャートが示されている。図１１に示すように、ノード装置１００ａのプロセッサ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、ローカルメモリであるＭＵ３０ａであることを把握する。ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理をＳＩＵ４５ａに行なわせ、これと並行して投機リードパス５００によって、ＭＵ３０ａへのリード制御をＭＩＵ３５ａに指示する（図１１の符号Ａ）。投機読み出しは、「ＰＩＵ４０ａ→ＭＩＵ３５ａ→ＭＵ３０ａ→ＭＩＵ３５ａ→ＰＩＵ４０ａ」なる経路で、ローカルメモリであるＭＵ３５ａからデータを読み出すことによって行なわれる。一方、ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理を行うためにＳＩＵ４５ａに対してトランザクションを発行する。ＳＩＵ４５ａは、タグメモリ３１０に対してスヌープ処理を行うためトランザクションをＩＣＮ３００経由で発行する。すると、ルーティング部３２０の動作によって、トランザクションは、「ＳＩＵ４５ａ→ＩＣＮ３００→タグメモリ３１０→ＩＣＮ３００→ＳＩＵ４５ａ」の経路で実行され、タグメモリ３１０からスヌープ情報を読み出す。ＳＩＵ４５ａは、ステータス情報が「Ｕ」であるので、このトランザクションのスヌープ結果が「Modify無し」（即ち、いずれのキャシュメモリにも存在しない）であると判断し、ＰＩＵ４０ａに対してその旨の判定情報を送る。

【００３３】

ＰＩＵ４０ａでは、プロセッサ１０ａから発行されたリード系トランザクションをＭＩＵ３５ａ、ＳＩＵ４５ａに発行済みであること（ＴＶビット）、ＭＵ３０ａからＭＩＵ３５ａ経由でローカルメモリのデータが返却されていること（ＬＶビット）、ＳＩＵ１５ａからスヌープ結果としてトランザクションでアクセスするラインがリモートノード１００ｂのキャシュメモリ２０ｂに「Modify」として登録されていないことを示す「Modify無し」の情報が報告されたこと、の３つ

の条件が満たされたことにより本トランザクションのリプライデータを、ローカルメモリ（ＭＵ３０ａ）から読み出したデータであるとしてプロセッサ１０ａにリプライデータを送る。

【００３４】

次に、上記リード系トランザクションにおける動作を、図５～図１０等を参照しつつより詳細について説明する。ＣＰＵ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａは、投機リードパス５００を用いＭＩＵ３５ａ経由でローカルメモリであるＭＵ３０ａに対してリード系トランザクションを発行する。同時に、タグメモリ３１０に対するスヌープ処理を行うために、ＳＩＵ４５ａに対してリード系トランザクションを発行する。ＰＩＵ４０ａはＭＩＵ３５ａ、ＳＩＵ４５ａに対してトランザクションを発行したため、ＲＤＣＢ４１０の対応する制御ユニットのTVビットを立てる。ＭＵ３０ａに対して発行したトランザクションは、そのリプライデータをＭＩＵ３５ａ経由でＰＩＵ４０ａに送る。ＰＩＵ４０ａは、送られてきたリプライデータをＲＤＢ４２０の対応するエントリに登録すると共に、ＲＤＢ４２０に登録したこのエントリに対するＲＤＣＢ４１０の制御ユニットのLVビットを立てる。

【００３５】

一方、ＳＩＵ１５ａに対して発行したトランザクションは、タグメモリ３１０に対するスヌープ処理の結果が返却されるまで、ＳＣＢ４５０内の対応するエントリのVビットを立てる。ローカルメモリの読み出しと並行して、タグメモリ３１０に対するスヌープ処理を行うため、「ＩＣＮ３００→タグメモリ３１０→ＩＣＮ３００」の経路でルーチング部がタグ情報を読み出し、読み出したタグ情報をＳＩＵ４５ａに送る。送られてきたタグ情報は、ＳＣＢ４５０内の対応するエントリのＳＮＢフィールドに登録される。そして、ＶビットとＳＮＢ情報とがそろうと、判定部４６０はトランザクションのスヌープ結果を判断情報として出力し、ＲＤＣＢ４１０内の対応する制御ユニットのＪＶビットに、この判断情報を登録させる。この例では、タグ情報が「Modify無し」となっているため、リプライデータとしては、ローカルメモリ（ＭＵ３０ａ）から読み出されたリードデータを採用すべきことが判断できる。ＰＩＵ４０ａにおいては、リード系トランザ

クションに対応したＲＤＣＢ４１０のエントリ情報として、ＴＶビット、ＬＶビット、ＪＶビットが全て立てられ、かつ、ＪＶビットは「Modify無し」であることから、Vビット生成回路４１２はEVビットを立てる（図１０の①参照）。EVビットが立ったエントリは、リプライ調停部４３０により調停出力され、これに対応するデータがＲＤＢ４２０からリプライデータとして出力され、リード系トランザクション発行元のプロセッサ１０ａへリプライデータを送ることでリード系トランザクションの処理が完了する。

【００３６】

したがって従来の処理によれば、図１１の下側図に示すように、まず、タグ情報を読み出し、これがステータス情報「U」であれば、今度は、ローカルメモリに対する読み出し動作を行なっていたので、本発明の実施形態の投機読み出しを行なうことにより、レイテンシの短縮が図られることが分かる。

【００３７】

（第２の動作例）

次に、図１０のケース②、即ち、ノード装置１００ａをホームノードとし、ローカルメモリ（ＭＵ３０ａ）に対するリード系トランザクションが、プロセッサ１０ａから発行された場合について、図１２等を参照しつつ説明する。なお、この時、図１６に示すようにタグメモリ３１０に記憶されるブロック番号ａのブロックデータに属するいずれかのデータをリード系トランザクションの読み出し対象とし、ノード装置１００ｂのノード装置番号を「②」とする。

【００３８】

図１２には、このケースの動作の処理手順を示すタイムチャートが示されている。図１１に示すように、ノード装置１００ａのプロセッサ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、ローカルメモリであるＭＵ３０ａであることを把握する。ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理をＳＩＵ４５ａに行なわせ、これと並行して投機リードパス５００によって、ＭＵ３０ａへのリード制御をＭＩＵ３５ａに指示する（図１１の符号Ｂ）。

投機読み出しは、「ＰＩＵ４０ａ→ＭＩＵ３５ａ→ＭＵ３０ａ→ＭＩＵ３５ａ→ＰＩＵ４０ａ」なる経路でローカルメモリであるＭＵ３５ａからデータを読み出すことによって行なわれる。一方、ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理を行うためにＳＩＵ４５ａに対してトランザクションを発行する。ＳＩＵ４５ａは、タグメモリ３１０に対してスヌープ処理を行うためトランザクションをＩＣＮ３００経由で発行する。すると、ルーティング部３２０の動作によって、トランザクションは、「ＳＩＵ４５ａ→ＩＣＮ３００→タグメモリ３１０→ＩＣＮ３００→ＳＩＵ４５ａ」の経路で実行され、タグメモリ３１０からスヌープ情報を読み出す。ＳＩＵ４５ａは、ステータス情報が「Ｐ」であるので、このトランザクションのスヌープ結果が「Modify有り」（即ち、ノード装置１００ｂのキャシュメモリに存在する）であると判断し、他ノード装置１００ｂのＳＩＵ４５ｂに対して、キャッシュデータ読み出しのための制御情報を送る。そして、「ＳＩＵ４５ｂ→ＰＩＵ４０ｂ→プロセッサ１０ｂ→キャッシュメモリ２０ｂ→プロセッサ１０ｂ→ＰＩＵ４０ｂ→ＳＩＵ４５ｂ→ＩＣＮ３００」なる経路でキャッシュデータがＩＣＮ３００に送られ、ルーティング部３２０は、これをノード装置１００ａのＳＩＵ４５ａに送る。すると、ＳＩＵ４５ａは、これをＰＩＵ４０ａに送る。

【００３９】

ＰＩＵ４０ａでは、プロセッサ１０ａから発行されたリード系トランザクションをＭＩＵ３５ａ、ＳＩＵ４５ａに発行済みであること（ＴＶビット）、ＭＵ３０ａからＭＩＵ３５ａ経由でローカルメモリのデータが返却されていること（ＬＶビット）、リモートノード１００ｂのキャッシュメモリ２０ｂに記憶するキャッシュデータが送られてきたこと（ＣＶビット）、ＳＩＵ１５ａからスヌープ結果としてトランザクションでアクセスするラインがリモートノード１００ｂのキャシュメモリ２０ｂに「Modify」として登録されていることを示す「Modify有り」の情報が報告されたこと、の４つの条件が満たされたことにより本トランザクションのリプライデータを、ノード装置１００ｂのキャッシュメモリ２０ｂから読み出したデータであるとしてプロセッサ１０ａにリプライデータを送る。なお、先にＭＵ３０ａから読み出されたデータは廃棄される。

【００４０】

このように第１および第２の動作例によれば、ノード装置１００ａは、タグ情報をタグメモリ３１０から読み出すのと並行して、自ノード装置１００ａ内のメモリユニット（ＭＵ）３０ａからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータ（例えば同一アドレスまたは同一アドレス範囲のデータ）が全キャッシュメモリ２０ａ、２０ｂにおいて存在しないことを示す場合には、この投機読み出しデータを自ノード装置１００ａ内のプロセッサ１０ａに送る。一方、読み出したタグ情報が、投機読み出し対象データと同一のデータがキャッシュメモリ２０ａまたは２０ｂにおいて存在することを示す場合には、このキャッシュメモリ２０ａまたは２０ｂに存在するデータを獲得して自ノード装置１００ａ内のプロセッサ１０ａに送り、投機読み出しデータを廃棄する。

【００４１】

したがって、従来のように（図１１上側図参照）、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データと同一のデータが全キャッシュメモリ１０ａ、１０ｂに存在しないことを示す場合には、さらに、自ノード装置１００ａ内のメモリユニット（ＭＵ）３０ａからのデータ読み出し動作を行なうといった一連の動作を行なうことはせず、ダグ情報の読み出しと並行して自ノード装置１００ａ内のメモリユニット（ＭＵ）３０ａから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータが全キャッシュメモリ２０ａ、２０ｂにおいて存在しないことを示す（例えばステータス情報「Ｕ」）場合には、既に投機読み出ししておいたデータをプロセッサ１０ａに送るので、レイテンシを短縮することができる。

【００４２】

（第３の動作例）

次に、図１０のケース③、即ち、他ノード装置１００ｂをホームノードとし、メモリユニット（ＭＵ３０ｂ）に対するリード系トランザクションが、プロセッサ１０ａから発行された場合について、図１３等を参照しつつ説明する。なお、この時、図１７に示すようにタグメモリ３１０に記憶されるブロック番号ｂのブ

ロックデータに属するいずれかのデータをリード系トランザクションの読み出し対象とする。

【００４３】

図１３には、このケースの動作の処理手順を示すタイムチャートが示されている。図１３に示すように、ノード装置１００ａのプロセッサ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、リモートノード１００ｂ内のＭＵ３０ｂであることを把握する。ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理をＳＩＵ４５ａに行なわせ、これと並行してＭＵ３０ｂに対する投機読み出し行なうために、ＩＣＮ３００に対して制御情報を送る。ルーティング部３２０は、これをＳＩＵ４５ｂに送る（図１３に符号Ｃ）。投機読み出しは、「ＳＩＵ４５ｂ→ＭＩＵ３５ｂ→ＭＵ３０ｂ→ＭＩＵ３５ｂ→ＳＩＵ４５ｂ」なる経路でリモートノード１００ｂのＭＵ３０ｂからデータを読み出すことによって行なわれる。そして、この読み出しデータはＩＣＮ３００に送られ、そのルーティング部３２０の動さによってノード装置１００ａのＳＩＵ４５ａを経由して、ＰＩＵ４０ａに送られる。

【００４４】

一方、ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理を行うためにＳＩＵ４５ａに対してトランザクションを発行する。ＳＩＵ４５ａは、タグメモリ３１０に対してスヌープ処理を行うためトランザクションをＩＣＮ３００経由で発行する。するとルーティング部３２０の動作によって、トランザクションは、「ＳＩＵ４５ａ→ＩＣＮ３００→タグメモリ３１０→ＩＣＮ３００→ＳＩＵ４５ａ」の経路で実行され、タグメモリ３１０からスヌープ情報を読み出す。ＳＩＵ４５ａは、ステータス情報が「Ｕ」であるので、このトランザクションのスヌープ結果が「Modify無し」（即ち、いずれのキャシュメモリにも存在しない）であると判断し、ＰＩＵ４０ａに対してその旨の判定情報を送る。

【００４５】

ＰＩＵ４０ａでは、プロセッサ１０ａから発行されたリード系トランザクショ

ンをＭＩＵ３５ａ、ＳＩＵ４５ａに発行済みであること（ＴＶビット）、他ノード装置のＭＵ３０ｂからデータが返却されていること（ＨＶビット）、ＳＩＵ１５ａからスヌープ結果としてトランザクションでアクセスするラインがリモートノード１００ｂのキャシュメモリ２０ｂに「Modify」として登録されていないことを示す「Modify無し」の情報が報告されたこと、の３つの条件が満たされたことにより本トランザクションのリプライデータを、ローカルノード１００ｂのＭＵ３０ｂから読み出したデータであるとしてプロセッサ１０ａにリプライデータを送る。

【００４６】

（第４の動作例）

次に、図１０のケース④、即ち、他ノード装置１００ｂをホームノードとし、メモリユニット（ＭＵ３０ｂ）に対するリード系トランザクションが、プロセッサ１０ａから発行された場合について、図１４等を参照しつつ説明する。なお、この時、図１８に示すようにタグメモリ３１０に記憶されるブロック番号ｂのブロックデータに属するいずれかのデータをリード系トランザクションの読み出し対象とし、ノード装置１００ｂのノード装置番号を「②」とする。

【００４７】

図１４には、このケースの動作の処理手順を示すタイムチャートが示されている。図１４に示すように、ノード装置１００ａのプロセッサ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、ローカルノード１００ｂのＭＵ３０ｂであることを把握する。ＰＩＵ４０ａは、タグメモリ３１０に対するスヌープ処理をＳＩＵ４５ａに行なわせ、これと並行してＭＵ３０ｂに対する投機読み出し行なうために、ＩＣＮ３００に対して制御情報を送る。ルーティング部３２０はこれをＳＩＵ４５ｂに送る（図１３に符号Ｄ）。投機読み出しは、「ＳＩＵ４５ｂ→ＭＩＵ３５ｂ→ＭＵ３０ｂ→ＭＩＵ３５ｂ→ＳＩＵ４５ｂ」なる経路でリモートノード１００ｂのＭＵ３０ｂからデータを読み出すことによって行なわれる。そして、この読み出しデータはＩＣＮ３００に送られ、そのルーティ

ング部３２０の動作によってノード装置１００ａのＳＩＵ４５ａを経由して、ＰＩＵ４０ａに送られる。

【００４８】

一方、ＳＩＵ４５ａは、ステータス情報が「Ｐ」であるので、このトランザクションのスヌープ結果が「Modify有り」（即ち、ノード装置１００ｂのキャシュメモリに存在する）であると判断し、他ノード装置１００ｂのＳＩＵ４５ｂに対して、キャッシュデータ読み出しのための制御情報を送る。そして、「ＳＩＵ４５ｂ→ＰＩＵ４０ｂ→プロセッサ１０ｂ→キャッシュメモリ２０ｂ→プロセッサ１０ｂ→ＰＩＵ４０ｂ→ＳＩＵ４５ｂ→ＩＣＮ３００」なる経路でキャッシュデータがＩＣＮ３００に送られ、ルーティング部３２０は、これをノード装置１００ａのＳＩＵ４５ａに送る。すると、ＳＩＵ４５ａは、これをＰＩＵ４０ａに送る。

【００４９】

ＰＩＵ４０ａでは、プロセッサ１０ａから発行されたリード系トランザクションをＭＩＵ３５ａ、ＳＩＵ４５ａに発行済みであること（ＴＶビット）、他ノード装置のＭＵ３０ｂからデータが返却されていること（ＨＶビット）、リモートノード１００ｂのキャッシュメモリ２０ｂに記憶するキャッシュデータが送られてきたこと（ＣＶビット）、ＳＩＵ１５ａからスヌープ結果としてトランザクションでアクセスするラインがリモートノード１００ｂのキャシュメモリ２０ｂに「Modify」として登録されていることを示す「Modify有り」の情報が報告されたこと、の４つの条件が満たされたことにより本トランザクションのリプライデータを、ノード装置１００ｂのキャッシュメモリ２０ｂから読み出したデータであるとしてプロセッサ１０ａにリプライデータを送る。なお、先にＭＵ３０ｂから読み出されたデータは廃棄される。

【００５０】

したがって、第３および第４の動作例によれば、ノード装置１００ａは、タグ情報をタグメモリ３１０から読み出すのと並行して、他ノード装置１００ｂ内のＭＵ３０ｂからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータ（例えば同一アドレスまたは同一ア

ドレス範囲のデータ）が全キャッシュメモリ２０ａ、２０ｂにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置１００ａ内のプロセッサ１０ａに送る。一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがキャッシュメモリ２０ａまたは２０ｂに存在することを示す場合には、このキャッシュメモリ２０ａまたは２０ｂに存在するデータを獲得して自ノード装置１００ａ内のプロセッサ１０ａに送り、投機読み出しデータを廃棄する。

【００５１】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データと同一のデータが全キャッシュメモリ２０ａ、２０ｂに存在しないことを示す場合には、さらに、他ノード装置１００ｂ内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、ダグ情報の読み出しと並行して他ノード装置１００ｂ内のＭＵ３０ｂから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータが全キャッシュメモリ２０ａ、２０ｂに存在しないことを示す（例えばステータス情報「Ｕ」）場合には、既に投機読み出しておいたデータをプロセッサ１０ａに送るので、レイテンシを短縮することができる。

【００５２】

また、以上説明してきた実施の形態においては、タグメモリ３１０を通信機構としてのＩＣＮ３００に設けているので、いずれのノード装置１００ａ、１００ｂからのタグ読み出しも略等しい時間で行なえるので、ノード装置間のレイテンシ差を小さくすることができる。

【００５３】

図１９は、本発明の他の実施形態のネットワークシステムのブロック構成図である。この実施形態においては、タグメモリ３１０をＩＣＮ３００に設けるのではなく、各ノード装置１００ａ、１００ｂ内に分割して設けた点に特徴があり、その他の点は図１に示した構成例と変わる所がない。図１９に示すように、ノード装置１００ａ、１００ｂには、夫々、タグメモリ３１１、３１２が設けられており、この両タグメモリ３１１、３１２のタグ情報を纏めたものが、上述してき

たタグメモリ３１０のタグ情報と一致する。

【００５４】

上述したケース①を例にとって、簡単に動作説明を行なう。プロセッサ１０ａからリード系トランザクションが発行されると、ＰＩＵ４０ａはＳＩＵ４５ａに対してスヌープ処理を実行させると共に、これと並行してＭＩＵ３５ａに対して投機読み出しを実行させる。投機読み出しは、「ＭＩＵ３５ａ→ＭＵ３０ａ→ＭＩＵ３５ａ→ＰＩＵ４０ａ」の経路で行なわれてＰＩＵは、ローカルメモリ（ＭＵ３０ａ）のデータを得てスヌープ待ち状態となる。この時点で、ＲＤＣＢ４１０の制御ユニットのＬＶビット、ＴＶビットが立つことになる。

【００５５】

一方、ＳＩＵ４５ａは、２つのタグメモリ３１１、３１２をスヌープする必要がある。したがって、タグメモリ３１１のステータス情報を獲得すると共に、ＩＣＮ３００に対してスヌープ処理実行のための情報を送る。なお、ここまでの処理中に、図２０（ａ）に示すようにＳＣＢ４５０のＶビットが立つ。次に、スヌープ処理実行のための情報は、ルーティング部３２０によって、ノード装置１００ｂのＳＩＵ４５ｂに送られ、ＳＩＵ４５ｂは、タグメモリ３１２のスヌープ処理結果をＩＣＮ３２０に送る。ルーティング部３２０は、これをノード装置１００ａのＳＩＵ４５ａに送る。したがって、図２０（ｂ）に示すようにＶビットが立つと共に、タグメモリ３１１のスヌープ結果を示す情報であるＳＮＰ１とタグメモリ３１２のスヌープ結果を示す情報であるＳＮＰ２とが登録されるので、判定部４６０は、これらの情報に基づいた判定結果をＲＤＣＢ４１０に送る。したがって、ＬＶビット、ＴＶビット、ＪＶビットがビットメモリ４１１に登録されるのでＶビット生成部４１２は、ＥＶビットを立てる。リプライ調停部４３０は、ＥＶビットが立ったので、これに対応するＲＤＢ４２０内のエントリをプロセッサ１０ａに送ることになる。ステータス情報が「Ｕ」で、投機読み出し対象データと同一のデータがキャッシュメモリ２０ａ、２０ｂに存在しない場合を想定すると、先に投機読み出しされたデータがリプライデータとしてプロセッサ１０ａに送られることになる。

【００５６】

このように、タグメモリ３１０を集中管理せずに、分割してネットワークシステムに分散するようにしても、投機読み出しによるレイテンシの短縮という同様の効果が得られることになる。また、タグメモリ３１０を通信機構内に設けなくても良いので通信機構の構成が簡素化される。

【００５７】

なお、今まで説明した動作を実行するためのデータアクセス用プログラムをコンピュータ読み取り可能な記録媒体に記録しておいて、プロセッサ１０ａ（１０ｂ）がこれを読み取り実行することによって以上説明した動作を行なわせるようにすれば良い。このような記録媒体としてはＲＯＭ等の半導体記録媒体、ＦＤ、ＨＤ（ハードディスク）等の磁気記録媒体、ＣＤＲＯＭ、ＤＶＤＲＯＭ等の光記録媒体等が挙げられる。もちろん、このようなアクセス用プログラムを、例えばハードディスク等で実現したローカルメモル（ＭＵ３０ａ，３０ｂ）の所定領域に記録しておいて、プロセッサ１０ａ（１０ｂ）がこれを実行するようにしても良い。

【００５８】

以上本発明の実施の形態について説明したきたが、本発明の要旨を逸脱しない範囲内で上述した実施形態に対して種々の変更や変形を施すことは可能であり、例えば、ノード装置の台数を３台以上にすること、各ノード装置内におけるプロセッサの台数を２台以上にすること、キャッシュメモリをプロセッサの外部に設けること等の変形が考えられることは言うまでもない。

【発明の効果】

以上説明したきたように、本発明によれば、投機読み出しをすることによって、ノード装置のキャッシュメモリやローカルメモリに対するアクセス時のレイテンシを短縮することが可能になるという効果が得られる。

【００５９】

また、タグ情報を記憶するタグメモリを通信機構に備えるようにして、ノード間のレイテンシ差を小さくすることができるという効果も得られる。

【図面の簡単な説明】

【図１】

本発明の実施の形態であるネットワークシステムのブロック構成図である。

【図２】

タグメモリ３１０に記憶されるタグ情報の説明図である。

【図３】

ステータス情報の説明図である。

【図４】

共有メモリのメモリマッピング状態を示す説明図である。

【図５】

本発明の実施形態の主要部の構成図である。

【図６】

ＲＤＣＢ４１０内の制御ユニットの構成図である。

【図７】

ビットメモリ４１１に格納される各ビットの説明図である。

【図８】

ＳＣＢ４５０の構成図である。

【図９】

ＳＣＢ４５０において用いる制御情報の説明図である。

【図１０】

動作説明のための説明図である。

【図１１】

第１の動作例と従来技術とのレイテンシの差を説明するためのタイミングチャートである。

【図１２】

第２の動作例を説明するためのタイミングチャートである。

【図１３】

第３の動作例を説明するためのタイミングチャートである。

【図１４】

第４の動作例を説明するためのタイミングチャートである。

【図１５】

第１の動作例を説明するためのタグ情報の具体例の説明図である。

【図１６】

第２の動作例を説明するためのタグ情報の具体例の説明図である。

【図１７】

第３の動作例を説明するためのタグ情報の具体例の説明図である。

【図１８】

第４の動作例を説明するためのタグ情報の具体例の説明図である。

【図１９】

本発明の他の実施の形態のネットワークシステムのブロック構成図である。

【図２０】

他の実施の形態のネットワークシステムの動作を説明するための説明図である。

【図２１】

従来技術の説明図である。

【図２２】

従来技術の説明図である。

【符号の説明】

１００ａ、１００ｂ　ノード装置

１０ａ、１０ｂ　プロセッサ

２０ａ、２０ｂ　キャッシュメモリ

３０ａ、３０ｂ　ＭＵ（メモリユニット）

３５ａ、３５ｂ　ＭＩＵ（メモリインターフェイスユニット）

４０ａ、４０ｂ　ＰＩＵ（プロセッサインターフェイスユニット）

４５ａ、４５ｂ　ＳＩＵ（システムインターフェイスユニット）

２００ａ、２００ｂ　ＳＣＵ（システムコントロールユニット）

３００　ＩＣＮ（インターナルコネクションネットワーク）

３１０　タグメモリ

３１１　タグメモリ

３１２　タグメモリ

３２０　ルーティング部

４１０　ＲＤＣＢ（リプライデータコントロールバッファ）

４１１　ビットメモリ

４１２　Ｖビット生成部

４２０　ＲＤＢ（リプライデータバッファ）

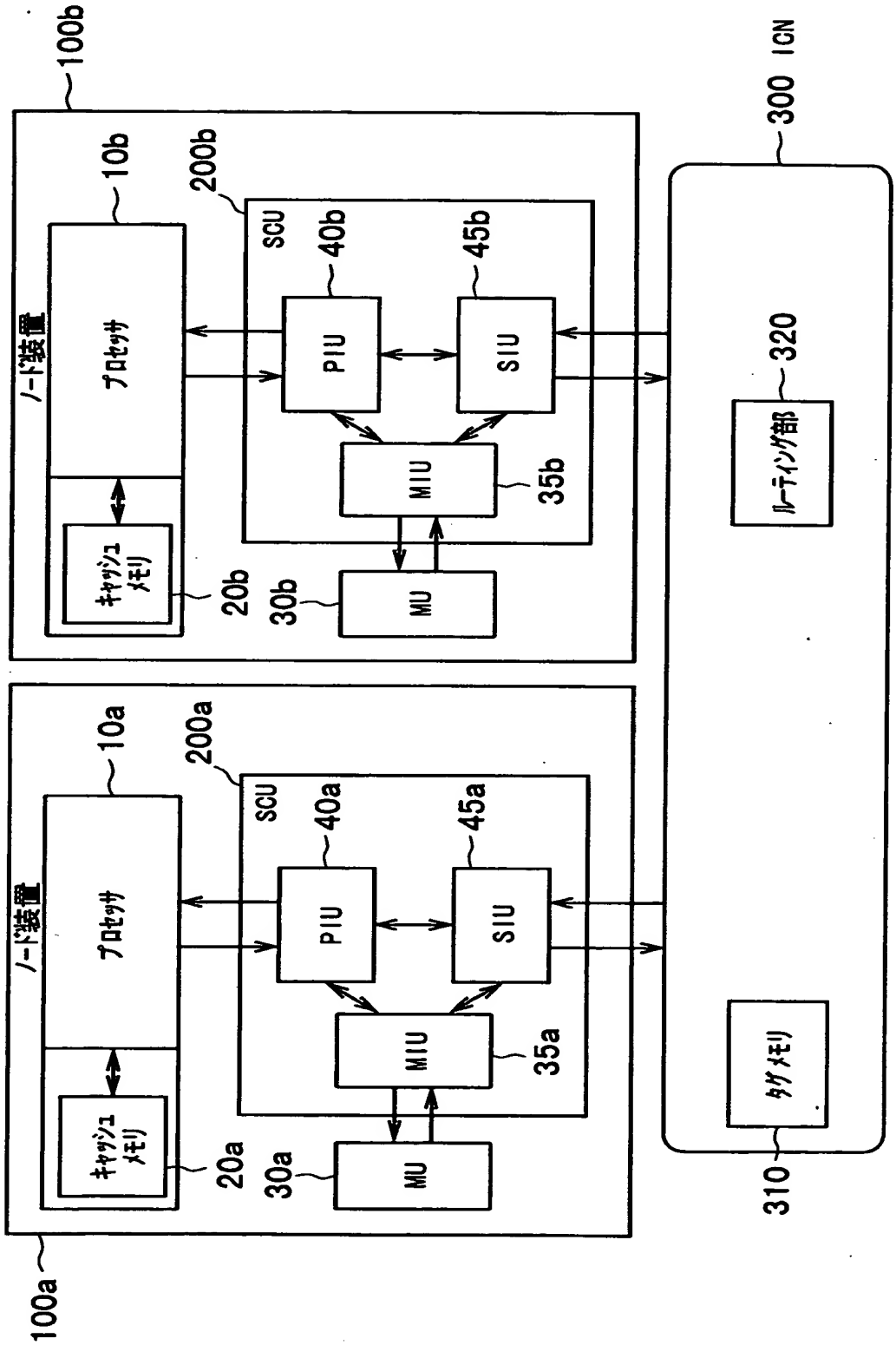
４３０　リプライ調停部

４５０　ＳＣＢ（スヌープコントロールバッファ）

４６０　判定部

【書類名】　図面

【図１】


100b
300 ICN
10b
200b
ノード装置
プロセッサ
SCU
40b
45b
PIU
SIU
320
ルーティング部
MIU
35b
キャッシュメモリ
20b
30b
MU
10a
200a
SCU
40a
45a
PIU
SIU
MIU
35a
キャッシュメモリ
20a
30a
MU
メインメモリ
310
100a

【図２】

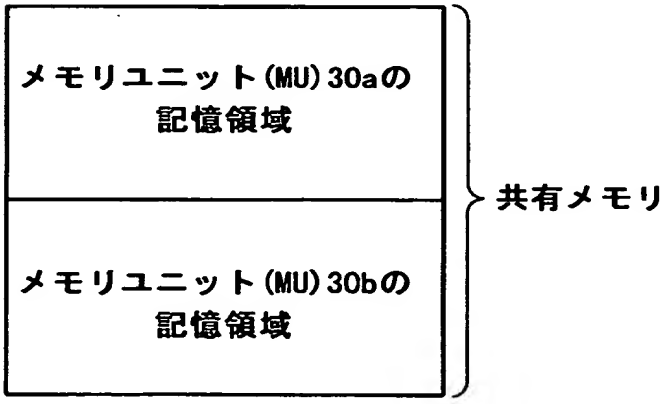
タグ情報

ブロック番号	ステータス情報	ノード装置番号
a	U	―――
b	S	１，２
⋮	⋮	⋮
n	P	２

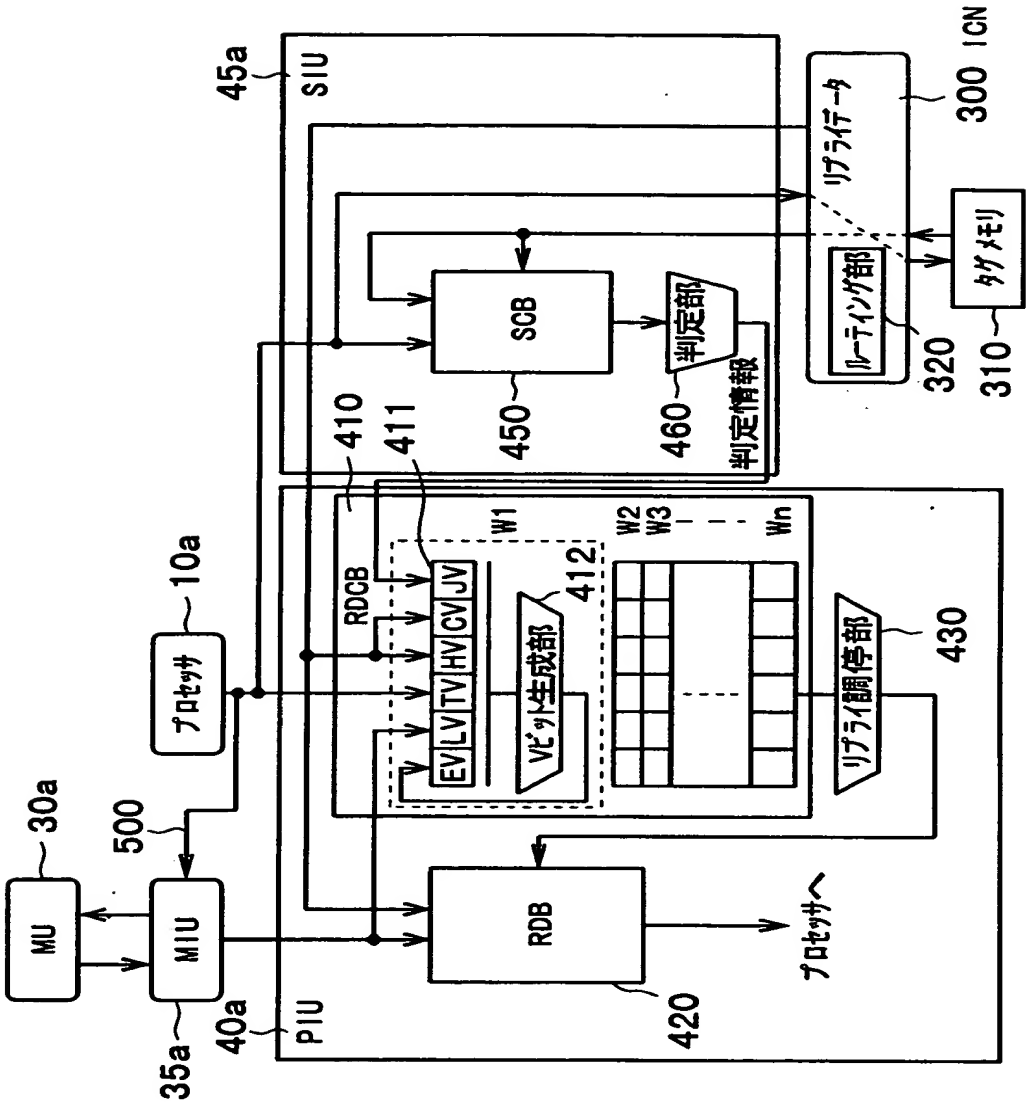
【図３】

ステータス情報	キャッシュ状態
U	いずれのノード装置においても、キャッシュデータが無い
S	キャッシュ内のデータは、メモリユニット内のデータと一致し、かつ、複数のノード装置において同一キャッシュデータ有り
P	１つのノード装置にのみ、キャッシュデータが有る

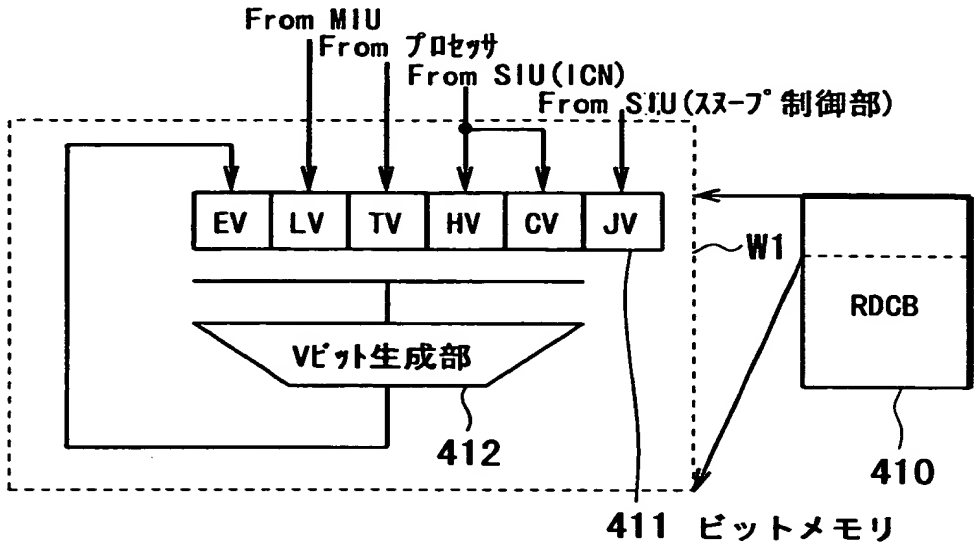
【図４】



【図５】


45a
SIU
300 ICN
リプライデータ
ルーティング部
320
タグメモリ
310
SCB
判定部
判定情報
450
460
410
411
W1
W2
W3
Wn
RDCB
EV
LV
TV
HV
CV
JV
412
Vビット生成部
リプライ調停部
430
10a
プロセッサ
30a
500
MU
MIU
35a
40a
PIU
RDB
420
プロセッサへ

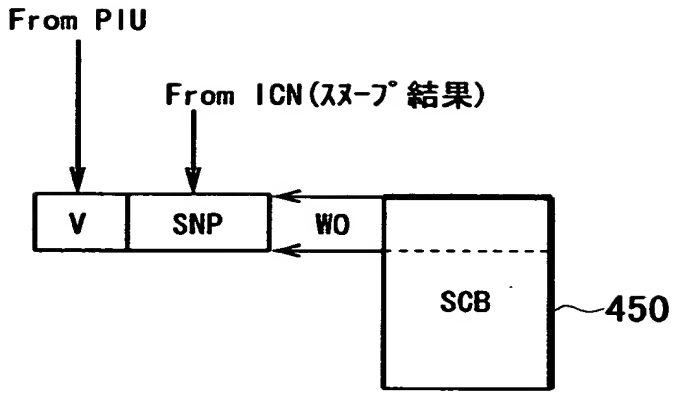
【図６】



【図７】

ＥＶ	プロセッサにリプライデータを送る条件が成立したことを示すエントリ代表ビット
ＬＶ	ローカルメモリから投機リードパスによるリプライデータが返却されたことを示すビット
ＴＶ	プロセッサからリード系トランザクションを受け付け、SIU/MIUに発行したことを示すビット
ＨＶ	他ノードに実装するメモリユニット(MU)からリプライデータが返却されたことを示すビット
ＣＶ	リモートノードのキャッシュに持つModifyデータが返却されたことを示すビット
ＪＶ	スヌープ結果を示すビット

【図８】



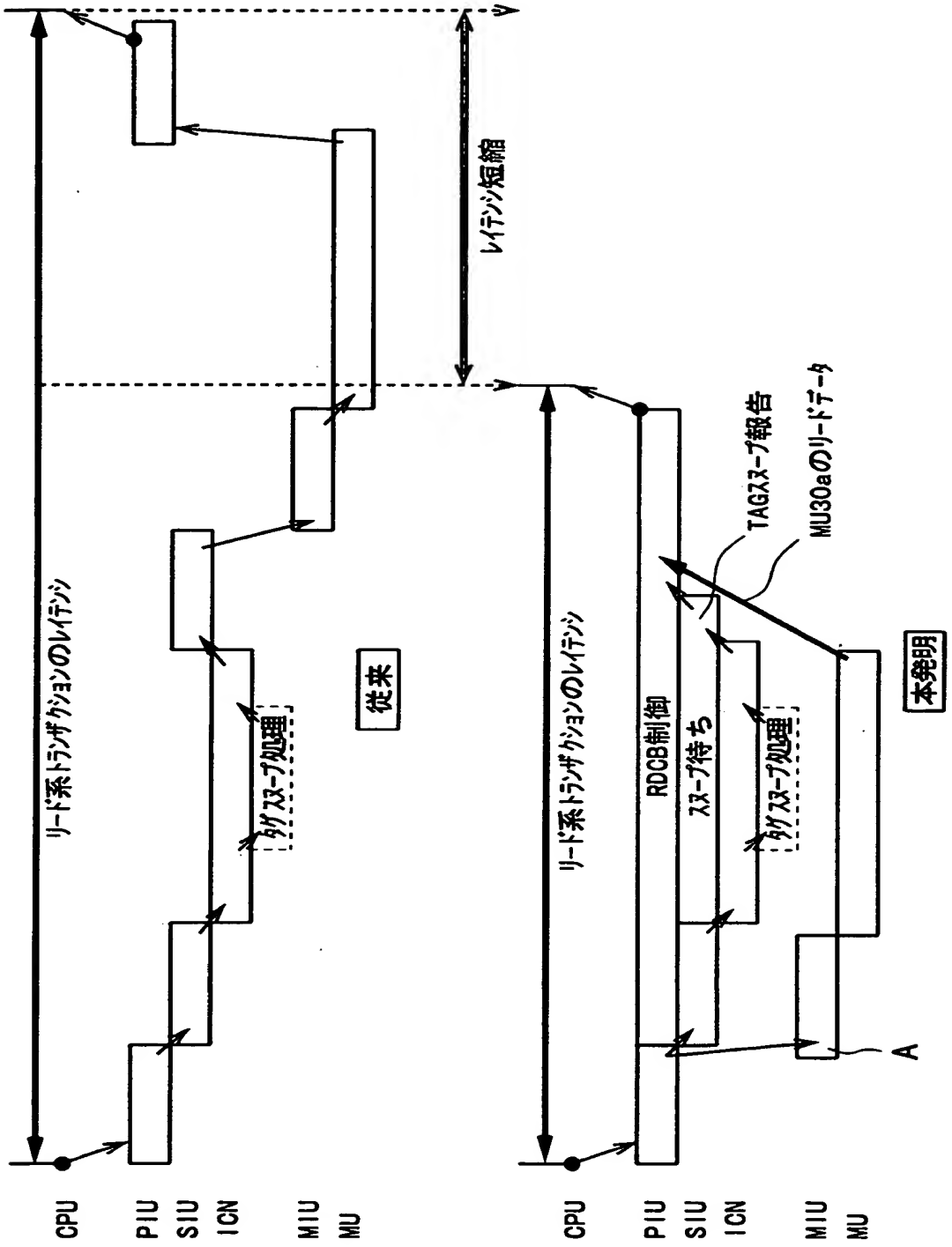
【図９】

V	プロセッサからリード系トランザクションが発行され、スヌープ処理待ちであることを示すビット
ＳＮＰ	Internal Connection Networkからのスヌープ結果を示す情報

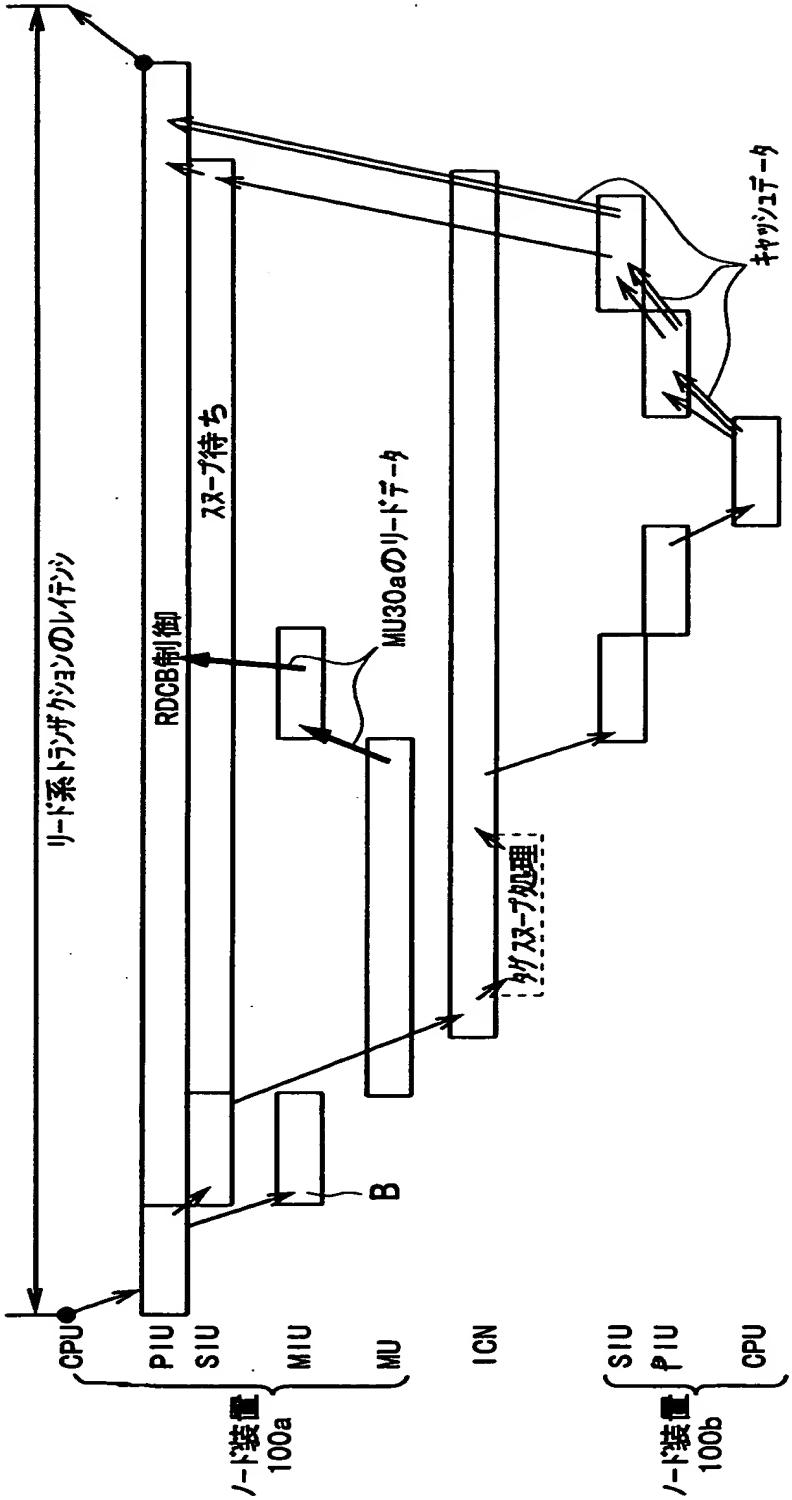
【図１０】

ｹｰｽ	LV	TV	HV	CV	JV	処理内容
①	1	1	0	0	Modify無	ローカルメモリデータをプロセッサに送る
②	1	1	0	1	Modify有	キャッシュデータをプロセッサに送り、ローカルメモリデータ(投機読み出しデータ)を廃棄する
③	0	1	1	0	Modify無	リモートノードのメモリデータをプロセッサに送る
④	0	1	1	1	Modify有	キャッシュデータをプロセッサに送り、リモートノードのメモリデータを廃棄する

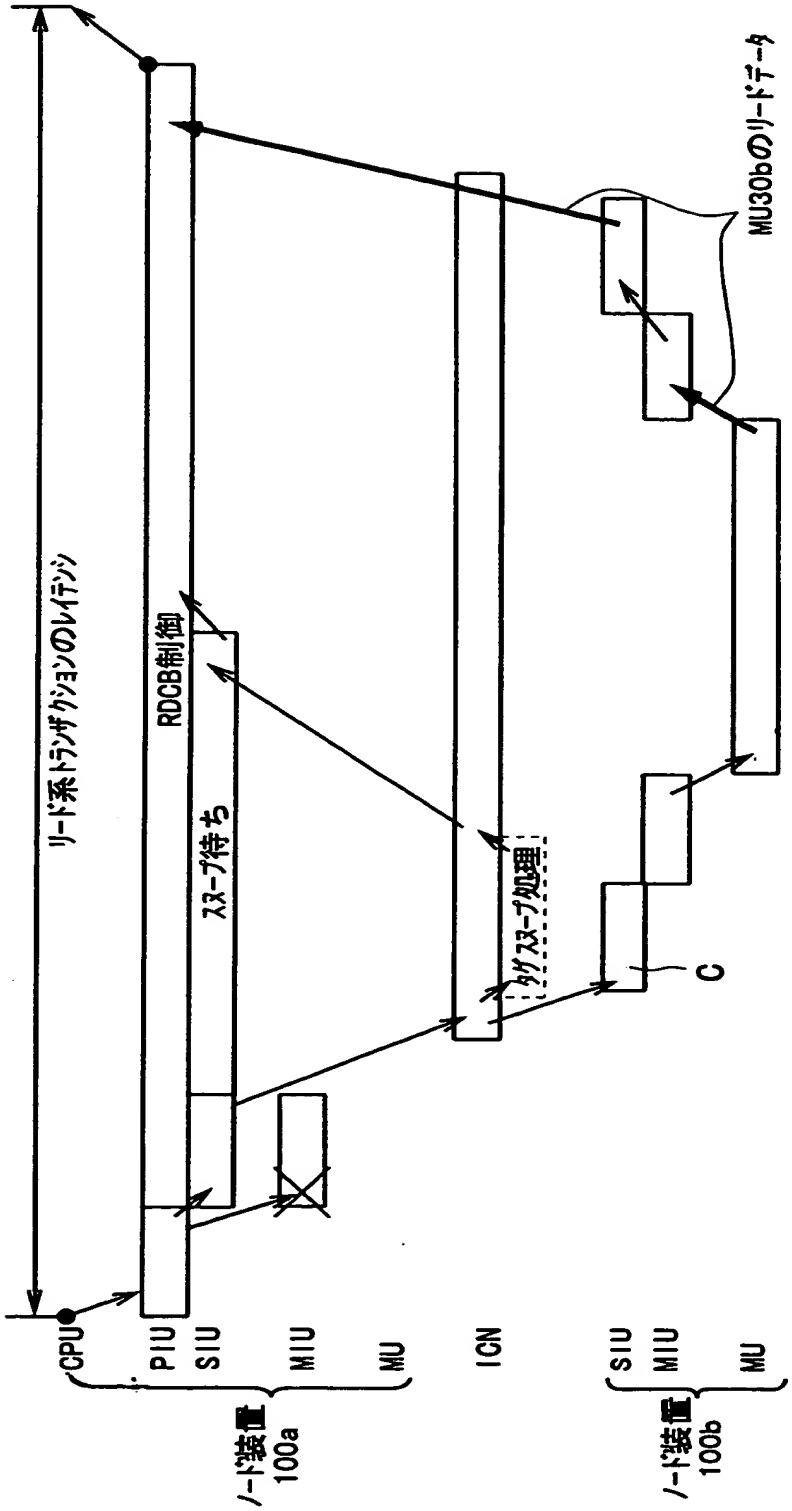
【図１１】


リード系トランザクションのレイテンシ
レイテンシ短縮
従来
外スヌープ処理
CPU
PIU
SIU
ICN
MIU
MU
リード系トランザクションのレイテンシ
RDCB制御
スヌープ待ち
TAGスヌープ報告
MU30aのリードデータ
本発明
外スヌープ処理
A
CPU
PIU
SIU
ICN
MIU
MU

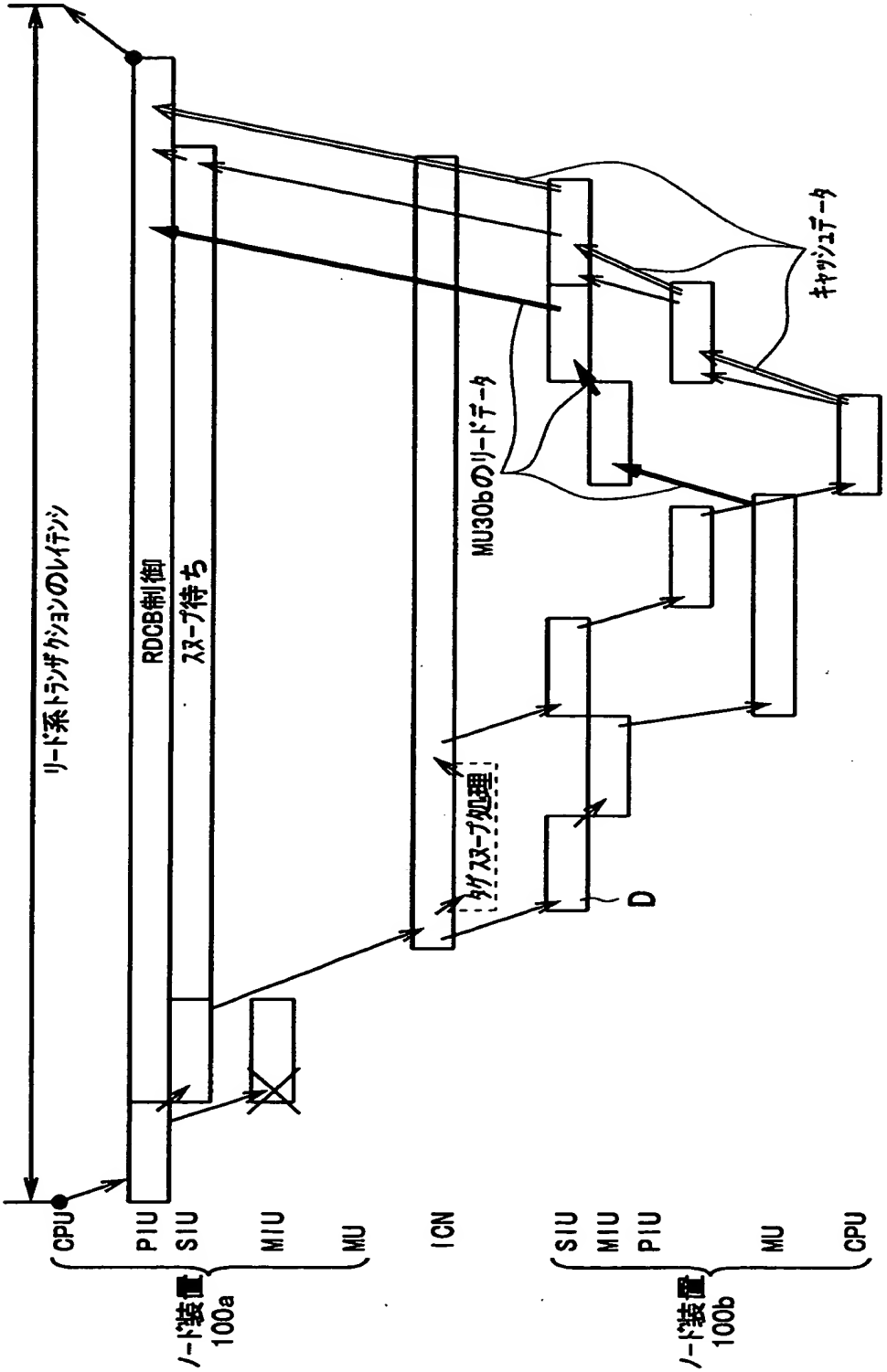
【図１２】


リード系トランザクションのレイテンシ
RDCB制御
スヌープ待ち
MU30aのリードデータ
キャッシュデータ
ライトスヌープ処理
B
CPU
PIU
SIU
MIU
MU
ICN
SIU
PIU
CPU
ノード装置 100a
ノード装置 100b

【図１３】


リード系トランザクションのレイテンシ
RDCB制御
スヌープ待ち
カノスヌープ処理
C
MU30bのリードデータ
CPU
PIU
SIU
MIU
MU
ICN
SIU
MIU
MU
リード装置
100a
リード装置
100b

【図１４】


リード系トランザクションのレイテンシ
RDOB制御
スヌープ待ち
MU30bのリードデータ
キャッシュデータ
タグスヌープ処理
D
CPU
PIU
SIU
MIU
MU
ICN
SIU
MIU
PIU
MU
CPU
ノード装置
100a
ノード装置
100b

【図１５】

ブロック番号	ステータス情報	ノード装置番号
a	U	―――
⋮	⋮	⋮

【図１６】

ブロック番号	ステータス情報	ノード装置番号
a	P	②
⋮	⋮	⋮

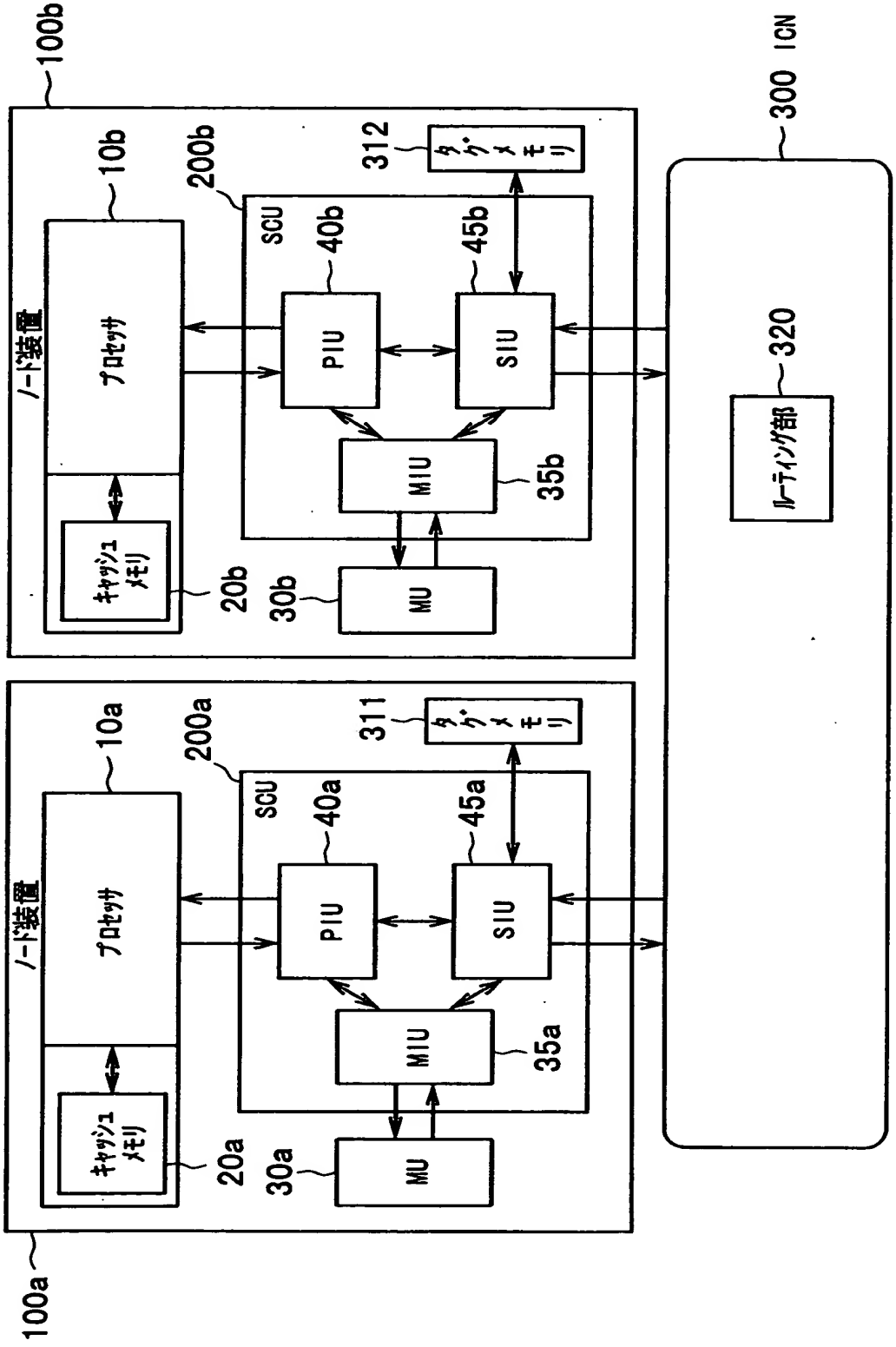
【図１７】

ブロック番号	ステータス情報	ノード装置番号
b	U	―――
⋮	⋮	⋮

【図１８】

ブロック番号	ステータス情報	ノード装置番号
b	P	②
⋮	⋮	⋮

【図１９】


100b
300 ICN
10b
200b
312
タグメモリ
SCU
40b
45b
ノード装置
プロセッサ
PIU
SIU
320
ルーティング部
キャッシュメモリ
MIU
35b
20b
30b
MU
10a
200a
311
タグメモリ
SCU
40a
45a
ノード装置
プロセッサ
PIU
SIU
キャッシュメモリ
MIU
35a
20a
30a
MU
100a

【図２０】

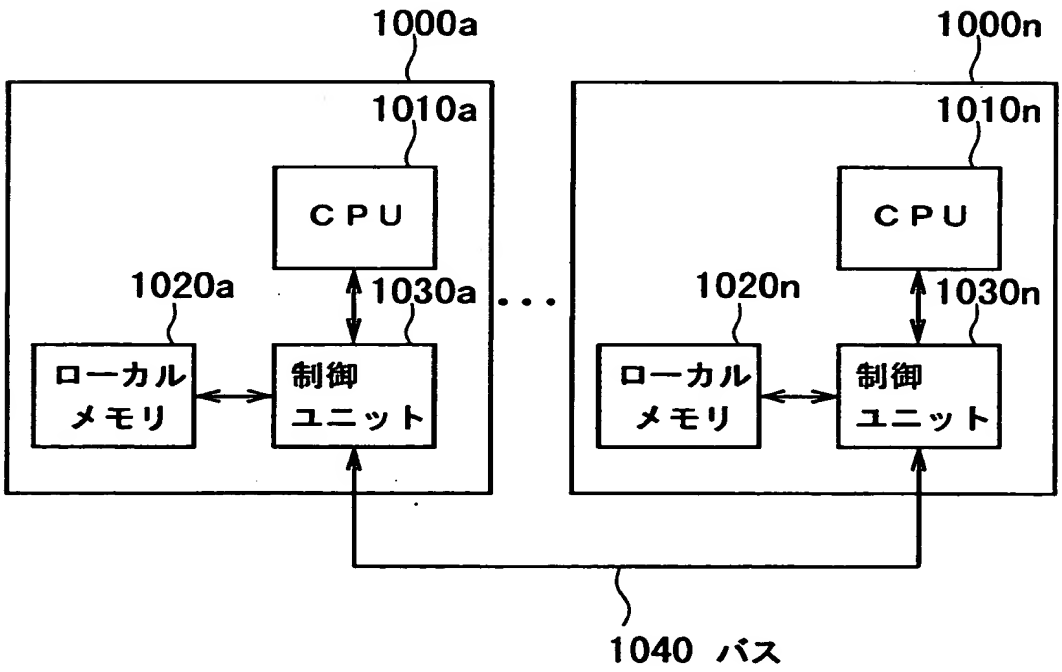
（ａ）

V	SNP1	SNP2
1		

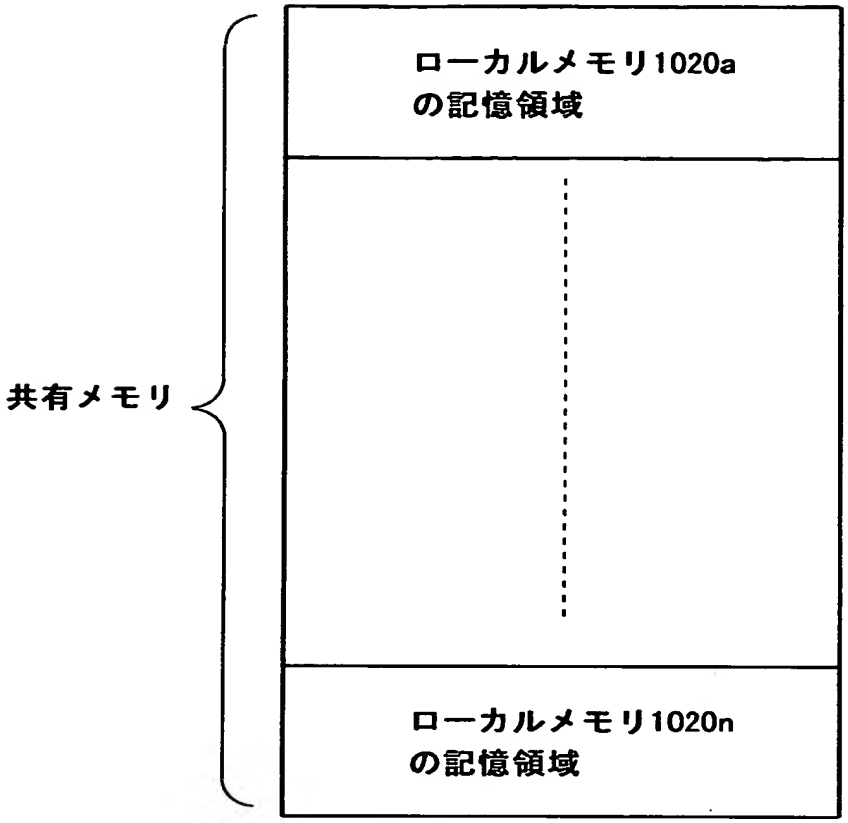
（ｂ）

V	SNP1	SNP2
1	スヌープ結果	スヌープ結果

【図２１】



【図２２】



【書類名】　要約書

【要約】

【課題】レイテンシを短縮できるデータアクセス方法を提供する。

【解決手段】タグ情報をタグメモリ３１０から読み出すのと並行して、自ノード装置１００ａ内のメモリユニット（ＭＵ）３０ａからデータを投機読み出しする。そして、読み出したタグ情報が、投機的読み出しデータが全キャッシュメモリ２０ａ、２０ｂにおいて存在しないことを示す場合には、この投機読み出しデータを自ノード装置１００ａ内のプロセッサ１０ａに送る。一方、読み出したタグ情報が、投機読み出しデータがキャッシュメモリ２０ａまたは２０ｂにおいて存在することを示す場合には、このキャッシュメモリ２０ａまたは２０ｂに存在するデータを獲得して自ノード装置１００ａ内のプロセッサ１０ａに送り、投機読み出しデータを廃棄する。

【選択図】　図１

出 願 人 履 歴 情 報

識別番号　〔０００１６８２８５〕

１．変更年月日　１９９０年　８月　９日
［変更理由］　新規登録
住　所　山梨県甲府市大津町１０８８－３
氏　名　甲府日本電気株式会社